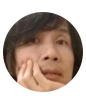
Face Recognition Is Robust With Incongruent Image Resolution: Relationship to Security Video Images

Article *in* Journal of Experimental Psychology Applied · April 2003

DOI: 10.1037/1076-898X.9.1.33 · Source: PubMed



Chang Hong Liu
Bournemouth University

Journal of Experimental Psychology: Applied
2003, Vol. 9, No. 1, 33–41

1076-898X/03/$12.00 DOI: 10.1037/1076-898X.9.1.33

Face Recognition Is Robust With Incongruent Image Resolution: Relationship to Security Video Images

Chang Hong Liu
McGill University

Helge Seetzen
University of British Columbia

A. Mike Burton
University of Glasgow

Avi Chaudhuri
McGill University

Identifying a criminal captured on conventional security video typically requires matching poor-quality video footage against a high-quality photograph. The authors examined the consequence of such a large discrepancy in image quality. Recognition and matching performance of this incongruent-quality condition was compared with that of a congruent one, in which a high-quality photograph was reduced to a low-quality video. Recognition memory was little affected by this manipulation, whereas matching performance of the incongruent condition enjoyed occasional advantage. The results show that person identification can tolerate a large discrepancy between image qualities of matching stimuli when one of the images is of poor quality.

Recent studies have shown that unfamiliar faces captured on commercially available video devices were identified poorly in both recognition and matching tasks, even though familiar faces were identified at near ceiling (Bruce, Henderson, Newman, & Burton, 2001; Burton, Wilson, Cowan, & Bruce, 1999). The lack of prior experience in forensic identification is unlikely to be a major factor for the poor performance because even highly experienced police officers performed no better on the same task than naïve participants (Burton et al., 1999). Because a common use of surveillance systems is to detect intruders or criminals who are often unfamiliar or unknown, these findings raise serious doubts about the reliability of images obtained from currently available surveillance systems.

The difficulty in recognizing unfamiliar faces from any sort of media is well-known. In a recent review, Hancock, Bruce, and Burton (2000) summarized how recognition of unfamiliar faces can be easily disrupted by variations such as lighting, viewpoint, facial expression, external features, and image quality, even though recognition of familiar faces is often robust and effortless under equivalent conditions. Thus, invariant recognition is only found for familiar faces. For unfamiliar faces, recognition is likely to depend on superficial similarities between face images. This has been characterized as picture recognition or image matching rather than as face recognition (Bruce & Young, 1986; Hancock et al., 2000). A change in lighting or facial pose can cause changes in luminance values of the corresponding pixels in the test image. However, the structural information of the face remains invariant after such transformation. Unfamiliar face recognition is highly susceptible to changes in low-level pictorial features, even though it may still be affected by high-level structural information (Liu & Chaudhuri, 2002; Liu, Collin, Rainville, & Chaudhuri, 2000).

A wide range of effects associated with recognition of unfamiliar faces may thus be traced to certain low-level pictorial determinants. One possible cause of the poor performance with closed-circuit television (CCTV) images may be due to discrepant image quality between the poor-resolution video at learning and the high-resolution photographs at test. This amounts to representing the same face at different spatial scales and therefore creates a considerable pictorial difference between the two images.

The question therefore arises as to whether a discrepancy in resolution between images for matching can produce a detrimental effect on face identification. This question may be approached from a low-level analysis of what is implied by seeing faces in different spatial resolution. Naturally, the facial information in a poor-quality video image is carried only by low-spatial frequencies, whereas a high-quality photograph contains facial information of both low- and high-spatial frequencies. Although a severely degraded image created by blurring can make it difficult to identify even familiar faces, recognition in such conditions shows strong resistance to degradation (Harmon, 1971). This is not surprising because if a familiar face in memory preserves both low- and high-spatial frequency information (Biederman & Kalocsai, 1997), then recognition of the familiar face in a blurred picture may be based on a match in the low-spatial frequency components alone.

Chang Hong Liu and Avi Chaudhuri, Department of Psychology, McGill University, Montreal, Quebec, Canada; Helge Seetzen, Department of Physics, University of British Columbia, Vancouver, British Columbia, Canada; A. Mike Burton, Department of Psychology, University of Glasgow, Glasgow, Scotland, United Kingdom.

A part of the data from this study was presented at the Vision Sciences Society Annual Meeting, Sarasota, Florida, May 2002. The study was supported by Medical Research Council Grant MOP-42514 and Natural Sciences and Engineering Research Council Grant 155482-98 to Avi Chaudhuri. We thank Elina Birmingham for helping to edit the video stimuli and for collecting some of the data in Experiment 1 and William H. A. Beaudot, Stéphane J. M. Rainville, Yi-Zhong Wang, Charles A. Collin, and Hugh Wilson for their suggestions on the design of stimuli.

Correspondence concerning this article should be addressed to Chang Hong Liu, who is now at the Department of Psychology, University of Hull, Cottingham Road, Hull HU6 7RX, United Kingdom. E-mail: c.h.liu@hull.ac.uk

Once a face in a low-quality image is recognized as a familiar person, a high-quality face image as a matching stimulus may supply further detail to verify whether the initial judgment was correct.

In contrast, when an unfamiliar face is shown in a low-quality image, the details in a high-quality matching stimulus may not help the process of identification because such details of the unknown face are absent in memory. An intriguing possibility is that the high-spatial frequency content of a photograph may actually interfere with the identification process. There is substantial evidence that unfamiliar face recognition is strongly influenced by spatial filtering (see a review by Costen, Parker, & Craw, 1996). That is, face stimuli containing a narrow band of frequencies are more difficult to recognize than broad or full-bandwidth images. Perhaps what is particularly relevant to the present study is the finding that low-pass filtered face images are more difficult to recognize than high-pass filtered ones when they are matched to unfiltered images (Fiorentini, Maffei, & Sandini, 1983). This finding shows that high-spatial frequency information about faces is useful, something that is lost in low-quality video images.

It appears that an important requirement for unfamiliar face recognition is how similar the two face images are. Identical images produce reasonably good recognition and matching performance, even when low-pass images contain frequencies as low as only 2.4 cycles per face (Liu et al., 2000). This does not necessarily mean that filtering has little effect as long as the images are identical. Recognition is always less accurate for severely filtered images than for less filtered images. However, the difference is much less dramatic than matching images that are filtered differently so that the two images are not identical. In this case, recognition is vulnerable to even a slight difference in the amount of spatial frequency overlap between matching stimuli. One possible cause of this effect is that facial information carried by the nonoverlapping spatial frequencies creates mismatches between the two stimuli.

On the basis of the sensitivity of unfamiliar face recognition to spatial frequency overlap and other overall image parameters, Liu et al. (2000) speculated that better performance should be possible if matching stimuli contain similar spatial frequencies. Also, recognition performance may be improved if a variety of image parameters, such as size, color, shading, and texture are made more similar between matching stimuli (Liu & Chaudhuri, 2000).

At present, little is known about whether the effects that arise from a difference in low-level representation can be offset by equating image similarity. Image manipulation within a single dimension such as spatial frequency content would help to isolate this factor from other similarity dimensions. However, such a reductionist approach does not necessarily warrant reliable predictions for complex realistic situations where multiple factors may interact. For example, although studies have shown a strong influence of spatial frequency overlap between matching stimuli, it is not clear whether such an influence on face recognition remains the same when the two images are taken under different lighting conditions. In such a manipulation, the two images would vary in more than one similarity dimension. Furthermore, it is not clear whether the effect of spatial frequency overlap can be applied to face identification with varying levels of resolution. Unlike spatially filtered images, both high- and low-resolution images contain broadband spatial frequencies.

To address these questions, we conducted several experiments using a video clip and a photograph of the same person taken at different times and under different settings. This resulted in two versions of the same face in different lighting, facial expression, background, and clothing. This is similar to a real situation where the available images of a suspect are more likely to be taken at different times and under different contexts. We conducted three experiments in which the quality of the video clips and photographic images was made to be either discrepant (incongruent) or similar (congruent). In the incongruent condition, high-quality photographs were used along with poor-quality video clips. In the congruent condition, high-quality photographs were reduced to the level of the video images. To optimize image similarity, the mean luminance and the size of the images in the congruent condition were also equalized. Both a standard old–new recognition task and a simultaneous-matching task were used. In these experiments, each person captured on video was either presented in four static snapshots (Experiment 1) or in a video clip (Experiment 2). Because the match or mismatch of external features may play an important role in face identification under both congruent and incongruent conditions, we also examined the effect by removing external features (Experiment 3).

General Method

Because there is a possibility that congruent and incongruent image quality may affect performance in recognition and matching tasks differently, participants in all experiments were asked to perform both tasks. The same video and photographic materials were used in both tasks. Participants always performed the recognition task before the matching task. We did not use the reverse order because face images presented during the matching task could later be confused with those appearing during the learning segment of the recognition task. A fixed order for the two tasks makes it difficult to compare their relative advantage but has little impact on our goals because we are not particularly interested in the relative merit of the two tasks. There is little doubt that the matching task is much easier than the recognition task. What we were mainly interested in was whether incongruent image quality has a similar effect on the two tasks.

We used a Power Macintosh G3 computer with an Apple Multiple Scan 20 Display to present stimuli and collect data. The screen resolution was set at 1152 $\times$ 870 pixels, with 32-bit colors. Face stimuli were shown on a neutral gray background that filled the entire screen. These stimuli are described later in the experimental sections. Programs for experimental control were written in MATLAB 5.1 for Macintosh, with Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

Design

The effect of congruity was assessed with a between-participants design. Results were analyzed using unpaired *t* tests. The quality of image resolution was either congruent or incongruent between the video image and the photographic image. The dependent variable was yes–no responses for the recognition task, and same–different responses for the matching task. The responses were used to compute d' that combined hit and false alarm. Criterion data were also used in the analysis.

Procedure

Participants were tested individually. The instructions along with an example of the stimuli were given on the monitor before the recognition and matching tasks. The details of each task were as follows:

Recognition task. A learning session with a presentation of video clips and a test session using photographic images as test stimuli were used. Participants were asked to decide whether the faces presented at the test session had been shown at the learning session. The stimuli were the same as in Burton et al. (1999), but a slightly different procedure was used in this study to avoid the close-to-floor performance implied in their rating data. In Burton et al.'s recognition task, 10 video clips were shown at learning session, and 20 high-quality photographs were shown at the test. We used only five faces in each learning session. Two separate learning sessions were given, each followed by a test session. To avoid confusing the faces in the two learning sessions, the learning faces were blocked by their gender, such that participants would learn either five female faces first or five male faces first. The order for the two blocks was randomized for each participant. The two blocks were separated by a short instruction on the screen, informing the participant that a keypress would start a new learning session for the faces in another gender.

The order of presentation for the five learning faces in each learning session was randomized for each participant. Each face was shown on the screen for 7 s. The five faces were shown twice. Participants were asked to press a key to learn the set one more time. The test session began immediately after the learning phase was completed. A set of 10 gender appropriate photographic images, 5 targets and 5 distractors, were presented one at a time in the center of the screen. To minimize a possible recency effect, the order of presentation of target faces during the testing session was the same as the second learning session, but distractors were randomly inserted into the sequence between targets. Participants were instructed to press the key labeled *yes* if the face was seen during the learning session, otherwise they were instructed to press the key labeled *no*. Participants were informed that the test images were different from the learned images but that some of the test images were the same faces as the learned ones. The test image remained on the screen until the participant responded.

The presentation for the second set of learning faces started immediately after the first test block was completed. The participant was informed that a new training session for faces in a different gender, initiated by a keypress, was about to start. The training and testing procedure for the second set of stimuli were the same as that for the first one.

As in Bruce et al. (2001) and Burton et al. (1999), the video images in this experiment contained the whole figure, whereas the test images (i.e., the photographs) contained only the heads. Burton et al. found that obscuring the body in video images produced no difference in recognition performance compared with fully visible versions, suggesting that observers mainly rely on facial information rather than on the shape of body and gait for their judgments. At the beginning of our experiments, participants were given examples of learning and test images and were reminded to pay attention to faces during the learning session because only faces were shown at the test session.

Matching task. This task immediately followed the recognition task. As with the recognition task, the face stimuli were presented in two blocks, divided by gender. At each trial, the video image was presented in the upper part of the screen; the photographic image was presented in the bottom part. The distance between the two images was 3 cm. Participants were instructed to judge whether the two images on the screen were of the same person. They were given as much time as needed to respond. For each face captured on the video, one target and one distractor of corresponding gender from photographic images were paired. The assignment of distractors to targets was completely random.

Because a different random order of presentation itself could contribute to the task performance, we generated one set of random orders, and each order was used twice—once for a participant in the congruent condition and once for a participant in the incongruent condition. Also, because different pairings of target and distractor stimuli in the matching task can itself produce different results, the same randomly assigned pair was used twice for 2 participants in the two conditions.

Experiment 1

This experiment examined the effect of resolution congruity on person identification using static snapshots from video clips. Bruce et al. (2001) showed that multiple stills (three in their study) captured from a video clip produced better performance than a single still. On the basis of this known advantage, we selected four stills from each video clip as target stimuli for this experiment. The four stills were presented simultaneously to participants.

Method

Participants. A total of 48 undergraduate students from McGill University, Montreal, Quebec, Canada, were evenly assigned to two conditions where the image resolution was either congruent or incongruent. The age of participants (19 women, 5 men) in the congruent group ranged from 18 to 38 years (Mdn = 20.0). The age in the incongruent group (21 women, 3 men) ranged from 18 to 31 years (Mdn = 20.5). All participants had normal or corrected-to-normal vision.

Materials. Both video and photographic images were created at the Department of Psychology at the University of Glasgow, United Kingdom. The video images were produced on a typical low-cost security surveillance system. The CCTV camera installed at the main entrance of a building turns on automatically each time a person enters or leaves the building, resulting in approximately 4 s of video recorded on a VHS tape. These video images were drawn from a subset of stimuli used in two previously published studies (Bruce et al., 2001; Burton et al., 1999), where more detailed descriptions can be found. Three sets of stimuli were edited from these original materials:

1. Video images: A total of 11 persons (6 men, 5 women) were used in this study. None was familiar to the participants in our experiment; however, one of them, a male, was used only as an example stimulus in the instructions. We used Adobe Premiere 5.0 to digitize and edit the video clips, eliminating a portion of the background that the persons never occupied. The resulting dimension of video was 230 × 430 pixels, which measured 7.7 cm × 14.5 cm on the computer monitor. Each video clip had a duration of 3.5 s. Four still images were then captured from each video clip. Two stills were selected around the starting and the end portions of each clip. The rest were selected around the two points that evenly divided the clip. An example of the resulting image is shown in Figure 1A. The layout of the four stills followed the temporal order of the video sequence. As Figure 1A shows, the four stills from left to right represented the sequence of a person at the door approaching in the direction of the CCTV camera.

2. High-resolution photographs: This stimulus set was composed of 21 digitized color images. The size of the image was 174 × 220 pixels (5.8

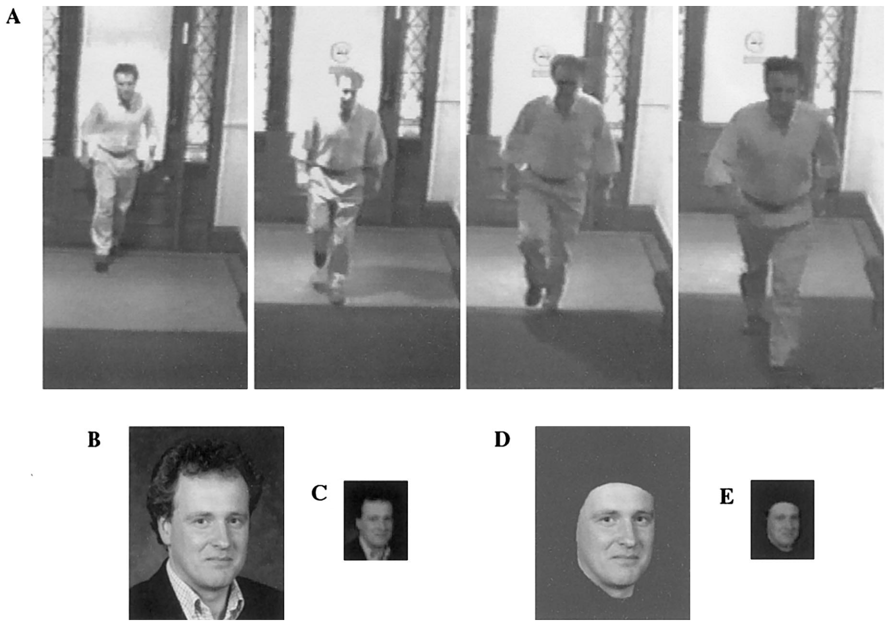


Figure 1. Example images of a target used in the study. A: Video images. In Experiments 1 and 3, four stills from the 3.5-s video clip were used. In Experiment 2, the video clip was played. B: High-resolution photograph. C: Low-resolution photograph. D: High-resolution photograph with external features removed. E: Low-resolution photograph with external features removed. B and C were used in Experiments 1 and 2; D and E were used in Experiment 3. The original stimuli in B and D were color images.

cm × 7.4 cm). Among these, 11 faces were the same as those in the video clips and served as target faces. One of the target faces was used only as an example during the instructions. The remaining 10 images served as distractors. Unlike persons in the video images, the bodies were not shown in the photographs. The photographs and videos were taken on different days and at different settings, causing variability in features such as clothing, hairstyle, and lighting. An example from this stimulus set is given in Figure 1B. This set of stimuli was used in the incongruent-resolution condition.

3. Low-resolution photographs: These images were created from the high-resolution photographs by lowering their resolution to the level of video images. The new low-resolution images were made to contain maximum resolution that the video still could provide. To achieve this, we first determined the resolution of the face images from the video stills. Because the four stills of each person were taken from different camera distances, the smaller face images resulted in lower resolution than the larger ones. To ensure that the maximum facial information provided in the video images could be used in our identification tasks, we based the estimation of resolution on the largest face in the four video stills. This corresponds to the rightmost still shown in the example of Figure 1A. To determine the resolution of a face, we took a row of pixels from one side of the ear to another (excluding the ears themselves) and counted the number of luminance chunks in the row. A luminance chunk consisted of adjacent pixels that had identical luminance values. The *resolution of face* was defined as the number of luminance chunks per face. The mean resolution of the largest faces was 22.5 luminance chunks ($SD = 5.8$). This mean was used to establish the resolution of the new images.

The low-resolution version of the photographs was made by the following procedure: First, the high-resolution color photographs were converted to 8-bit grayscale images. Then the images were clipped to a size of 72 × 90 pixels (2.4 cm × 3.0 cm), which approximated the size of the large faces in the video stimuli. The resulting images had an average resolution of 47.4 luminance chunks ($SD = 2.3$) per face, which was 2.1 times better than that of video images. To bridge this difference, we degraded the resized images by transforming each 2 × 2 block of pixels into a single brightness level. The average brightness value of the block was then assigned to fill the entire 2 × 2 block of pixels. After the blocking, a Gaussian blur was applied in both vertical and horizontal directions.

Apart from resolution, the luminance range of the generated images was also adjusted to that of the video images. We calculated the average grey value of the video stills by adding all the grey values of the image pixels and dividing it by the image resolution. This value was then used to adjust the brightness of the low-resolution photograph. The brightness was reduced or increased until it matched the video image. To match the low-contrast video images, the contrast of the newly created image was reduced by 40% using a simple contrast filter. An example of a resulting low-resolution photograph is shown in Figure 1C.

Results and Discussion

The means and their standard deviations for hits, correct rejections, and the combinations of the two are shown in Table 1. Statistical analyses were applied to d' and criterion data derived from the results of hit and false alarm rates. An alpha level of .05 was adopted for all analyses, and effect sizes for all comparisons were calculated using Cohen's d.

Unpaired t tests showed no significant difference between the d' results of congruent and incongruent conditions (Ms = 1.10 and 0.77, SDs = 0.80 and 0.74, respectively) in the recognition task, $t(46) = 1.44$, *ns*, $d = 0.42$. Furthermore, there was no significant difference between the d' of congruent and incongruent conditions (Ms = 2.38 and 2.52, SDs = 0.93 and 0.72, respectively) in matching task, $t(46) = -0.59$, *ns*, $d = 0.17$.

Response bias was comparable between congruent and incongruent conditions for the recognition task (Ms = 0.45 and 0.34, SDs = 0.42 and 0.40, respectively), $t(46) = 0.96$, *ns*, $d = 0.28$.

Table 1
Mean Percent Correct Responses and Standard Deviations as a Function of Image Congruity in Experiment 1

	Recognition task			Matching task		
Condition	Hit	CR	Overall	Hit	CR	Overall
Congruent						
M	53	79	66	62	87	75
SD	14	15	11	16	11	11
Incongruent						
M	52	74	63	73	85	79
SD	19	14	12	19	8	11

Note. CR = correct rejection.

However, for the matching task, response bias differed between the two conditions. The criterion scores showed that the participants in the congruent condition ($M = 0.50$, $SD = 0.40$) were more biased to answer "different" than those in the incongruent condition ($M = 0.19$, $SD = 0.38$), $t(46) = 2.73$, $p < .01$, $d = 0.81$.

These results indicate that person identification receives no benefit from congruent image quality. In fact, analyses on these tests revealed only moderate power (post hoc analysis with medium effect size, $d = 0.50$, reveal a moderate power of 0.53 for all of the comparisons above). However, the nonsignificant trend in the hit rate data from the matching task suggests congruent quality could well impair identification performance.

The overall poor performance in recognition memory of unfamiliar faces in comparison to similar tasks using high-quality images was consistent with Burton et al.'s (1999) finding. Our results showed that even after reducing the training set from 10 persons in the Burton et al.'s study to 5, and even after each set was shown twice, the overall percent correct response for the two groups could only reach 65%. This finding further demonstrates the vulnerability of recognition memory for unfamiliar persons.

The overall greater accuracy for the matching task clearly showed the effect of memory load. The combined percent accuracy for the congruent and incongruent conditions was 77%. At first sight, this may appear to be inconsistent with the result (66%) from a similar task in Bruce et al. (2001, Experiment 1). However, this apparent inconsistency was clearly due to the different respective methods of target–distractor pairing. In the Bruce et al. study, distractors (named *nontargets* in their article) were carefully selected to resemble the appearance of the targets. In our study, however, the pairing of targets and distractors was completely random, which only infrequently produced similar appearances. Thus, the pairing in the Bruce et al. study should have been more confusing for the observers than ours. This was reflected in the correct rejection results of the two studies. Their correct rejection was as low as 52%, whereas ours was 86%. The hit rate in Bruce et al. (79%) was actually higher than ours (68%) when results from congruent and incongruent conditions are combined. However, when the hit data from our congruent condition are disregarded, the mean percentage hit (73%) was remarkably close to that in Bruce et al. Such agreement, if statistically significant, could lend further support to our current finding that presenting a high-resolution photograph may facilitate the best identification performance even when the matching image is of poor quality.

Our criterion data showed that participants in both congruent and incongruent conditions were biased to respond negatively. The cautious bias in the incongruent condition is consistent with the Bruce et al. (2001) finding. For the incongruent condition, such bias is not surprising because a great image difference could result in negative responses. However, what appears to be paradoxical is the result that the congruent condition produced stronger negative bias than the incongruent condition; in other words, a greater similarity between the learned and test images prompted more "different" rather than "same" responses. This shift of criterion is associated with the impaired performance of the congruent condition.

To establish the generality of our finding in Experiment 1, we next tested the effect of image congruity on person identification from dynamic video clips.

Experiment 2

Studies on the role of dynamic information in face recognition have yielded conflicting results. Although Pike, Kemp, Towell, and Phillips (1997) showed that movement can help memory for unfamiliar faces, other studies were unable to confirm this benefit (Bruce et al., 1999, 2001; Christie & Bruce, 1998). Summarizing the evidence in the literature, O'Toole, Roark, and Abdi (2002) concluded that dynamic information is useful for identifying familiar faces, such as celebrities under poor-image resolution and illumination. However, whether dynamic information is helpful for recognition of unfamiliar faces remains unclear.

One of our goals in this study was to investigate whether dynamic information could benefit recognition of unfamiliar faces in the same way it does for familiar faces when poor-resolution images are presented. We were mainly interested in the role of image congruity in using dynamic information. Namely, can a photograph produce better identification when its image quality matches or is superior to that of the video clip? Does congruity in resolution affect face recognition with dynamic facial information in the same manner as it does with static face images? Participants were required to perform the same recognition and matching tasks as in Experiment 1, but the video clips were presented as movies rather than as a few static images.

Method

Participants. A total of 50 undergraduate students from McGill University participated in the study. Twenty-five of them (24 women, 1 man), whose age ranged from 18 to 28 years ($Mdn = 19$), participated in the condition where the image resolution was congruent. The other half (20 women, 5 men), whose age ranged from 18 to 34 years ($Mdn = 20$), were tested in the condition where the image resolution was incongruent. All participants had normal or corrected-to-normal vision.

Materials. The face stimuli were essentially the same as those used in Experiment 1, except that for the video images, the video clips rather than the still images were used in Experiment 2.

Procedure. A 3.5-s video clip for each target face was looped twice in a row at each learning trial during the training session of the recognition task. This length of exposure matched the 7-s presentation for the still images in Experiment 1. For the matching task, participants were allowed to press a key to view the same video clip as many times as they wanted. Other aspects of the experimental procedure were the same as those described in the General Method section.

Results and Discussion

Table 2 shows the percent correct responses for matched trials, unmatched trials, and the two combined. The sensitivity measure

Table 2
Mean Percent Correct Responses and Standard Deviations as a Function of Image Congruity in Experiment 2

	Recognition task			Matching task		
Condition	Hit	CR	Overall	Hit	CR	Overall
Congruent						
M	44	75	59	63	85	74
SD	14	12	9	17	9	8
Incongruent						
M	50	78	64	69	92	81
SD	17	12	9	21	9	10

Note. CR = correct rejection.

for the incongruent condition (M = 0.86, SD = 0.57) was not significantly better than for the congruent condition (M = 0.56, SD = 0.56) in the recognition task, $t(48) = 1.94$, *ns*, $d = 0.56$. However, there was an advantage to the incongruent condition ($M = 2.85$, $SD = 0.90$) than to the congruent condition ($M = 2.29$, $SD = 0.62$) in the matching task, $t(48) = 2.56$, $p < .01$, $d = 0.74$.

As with Experiment 1, no significant difference was found between the criterion results of congruent and incongruent conditions in the recognition task (*Ms* = 0.45 and 0.41, *SDs* = 0.29 and 0.40, respectively), $t(48) = 0.43$, *ns*, $d = 0.12$. However, unlike Experiment 1, the response bias of the congruent group ($M = 0.40$, $SD = 0.47$) for the matching task was no more cautious than that of the incongruent group ($M = 0.56$, $SD = 0.59$), $t(48) = -1.09$, *ns*, $d = 0.31$. All comparisons here have equivalent power (post hoc analysis with medium effect size, $d = 0.5$, reveal power of 0.54), allowing straightforward comparison of the significant and nonsignificant effects.

The d' results of the recognition task in Experiments 1 and 2 indicate that the congruent group performed better with still images rather than with dynamic images, whereas this manipulation had little effect on the incongruent group. This was confirmed by a 2 × 2 post hoc analysis of variance (ANOVA), which treated the stimulus type (dynamic vs. static) between the two experiments as one variable and congruity as another. Although neither the main effect of stimulus type nor congruity was significant, there was a significant interaction. The details of this analysis are shown in Table 3, along with the results of the same analysis on the matching data. There was no significant difference between the effects of dynamic and static stimuli in the matching task, but there was a significant advantage of the incongruent condition. Unlike the results of the recognition task, the interaction between stimulus type and congruity was not significant.

We also performed the same ANOVA tests on the criterion data. As Table 4 shows, the only significant effect of this analysis was the interaction between stimulus type and congruity in the matching task. The criterion of the congruent group showed little difference whether static or dynamic stimuli were used, whereas the criterion of the incongruent group became more conservative when dynamic stimuli were used.

Overall, the results showed that dynamic images had no advantage over static ones. For recognition memory, dynamic information seemed to hinder rather than facilitate identification when the image quality was congruent. However, such information had no harmful effects to either recognition or matching when the image quality was incongruent.

Table 3
Post Hoc ANOVAs for the Sensitivity Results in Experiments 1 and 2

Source	*df*	*MS*	*F*	Cohen's *f*
Recognition task				
Stimulus type (ST)	1	1.25	2.76	0.17
Congruity (C)	1	0.00	0.00	0.00
ST × C	1	2.42	5.34*	0.24
Error	94	(0.45)		
Matching task				
ST	1	0.35	0.55	0.08
C	1	2.99	4.66*	0.22
ST × C	1	1.06	1.66	0.13
Error	94	(0.64)		

Note. Values enclosed in parentheses represent mean square errors. ANOVAs = analyses of variance.
* $p < .05$.

Experiment 3

Person identification is known to rely on both internal and external features of a face (e.g., Campbell, Walker, & Baron-Cohen, 1995; Ellis, Shepherd, & Davies, 1979; Liu & Chaudhuri, 1998; Nachson, Moscovitch, & Umilta, 1995; Young, Hay, McWeeny, Flude, & Ellis, 1985). When faces have to be identified from poor-quality videos, the external features such as hair may play an even greater role than when the faces are identified from high-quality images. With high-quality images, facial details such as the fine shape of the eyes and mouth are readily available. Given that such useful information is not readily available in poor-quality images, observers may have to rely more on external features, such as hairstyle and the overall shape of the head. Poor-quality video images preserve this information because of the coarse scale of these features. As long as the external features are distinct between face stimuli, observers should be able to discriminate identity on the basis of external features alone. But when these features are similar, more false alarms should be predicted. This may be the cause of the poor performance found in the Bruce et al. study (2001, Experiment 1) in which distractors were selected to resemble the appearance of the targets.

Table 4
Post Hoc ANOVAs for the Criterion Results in Experiments 1 and 2

Source	*df*	*MS*	*F*	Cohen's *f*
Recognition task				
Stimulus type (ST)	1	0.04	0.27	0.05
Congruity (C)	1	0.15	1.03	0.10
ST × C	1	0.03	0.22	0.05
Error	94	(0.14)		
Matching task				
ST	1	0.43	1.98	0.15
C	1	0.13	0.58	0.08
ST × C	1	1.36	6.20**	0.26
Error	94	(0.22)		

Note. Values enclosed in parentheses represent mean square errors. ANOVAs = analyses of variance.
** $p < .01$.

Method

Participants. A total of 84 undergraduate students from McGill University were evenly assigned to two groups. The age of the participants (33 women, 9 men) in the congruent group ranged from 18 to 53 years (*Mdn* = 20). The age in the incongruent group (31 women, 11 men) ranged from 18 to 46 years (*Mdn* = 20). All participants had normal or corrected-to-normal vision.

Materials. External features of the faces in the photographic stimuli were excluded by manually removing all components outside the boundary delineated by the hairline and the lines between cheek and ear and between chin and neck. Figures 1D and 1E show an example for each of the high- and low-resolution photographs, respectively, after removing the external features. The video stimuli were identical to those used in Experiment 1.

Procedure. Participants were explicitly told that the photographic stimuli would not contain external features of faces and that they should pay attention to the internal features. The rest of the experimental procedure was identical to Experiment 1.

Results and Discussion

The means and standard deviations for hits, correct rejections, and the combinations of the two are shown in Table 5. For the recognition task, the d' results of congruent condition ($M = 0.47$, $SD = 0.66$) and incongruent condition ($M = 0.43$, $SD = 0.65$) were equivalent, $t(82) = 0.24$, *ns*, $d = 0.05$. Both groups performed poorly but were significantly better than chance, $ts(41) = 4.31$ and 4.59, respectively, $ps < .01$, $d = 1.35$ and 1.43, respectively. For the matching task, the d' was significantly higher in the incongruent condition ($M = 2.38$, $SD = 0.90$) than in the congruent condition ($M = 1.80$, $SD = 0.73$), $t(82) = 3.25$, $p < .01$, $d = 0.72$.

No significant difference was found between the criterion results of congruent and incongruent conditions in the recognition task ($Ms = 0.36$ and 0.26, $SDs = 0.39$ and 0.40, respectively), $t(82) = 1.23$, *ns*, $d = 0.27$. Unlike Experiment 1, but consistent with Experiment 2, the response bias of the congruent group ($M = 0.35$, $SD = 0.44$) for the matching task was comparable to that of the incongruent group ($M = 0.38$, $SD = 0.40$), $t(82) = -0.31$, *ns*, $d = 0.07$. All comparisons here have equivalent power (post hoc analysis with medium effect size, $d = 0.50$, reveal power of 0.73).

The mean d' results in Experiments 1 and 3 suggest that removing external features of faces clearly had a detrimental effect on the recognition task for both congruent and incongruent groups. This was confirmed by a post hoc ANOVA (Table 6). For the matching task, the results of the congruent group appeared to be more impaired than the incongruent group. However, this interaction was not significant. Both main effects were significant. The whole face stimuli were better identified than faces without external features, and the incongruent group outperformed the congruent group.

The same post hoc ANOVA applied to the criterion data only found a significant interaction between stimulus type and congruity in the matching task (Table 7). Faces with external features produced a more conservative criterion in the congruent group than in the incongruent group, whereas faces without any of these features produced little difference between the criterion results of the two groups.

Table 5
Mean Percent Correct Responses and Standard Deviations as a Function of Image Congruity in Experiment 3

Condition	Recognition task			Matching task		
	Hit	CR	Overall	Hit	CR	Overall
Congruent						
M	45	70	58	57	80	68
SD	17	15	10	21	13	10
Incongruent						
M	49	65	57	65	87	76
SD	16	16	11	19	9	11

Note. CR = correct rejection.

Table 6
Post Hoc ANOVAs for the Sensitivity Results in Experiments 1 and 3

Source	*df*	*MS*	*F*	Cohen's *f*
Recognition task				
Stimulus type (ST)	1	7.14	14.51**	0.34
Congruity (C)	1	0.97	1.97	0.13
ST × C	1	0.63	1.28	0.10
Error	128	(0.49)		
Matching task				
ST	1	3.98	5.87*	0.22
C	1	3.98	5.88*	0.22
ST × C	1	2.18	2.18	0.13
Error	128	(0.68)		

Note. Values enclosed in parentheses represent mean square errors. ANOVAs = analyses of variance.
* $p < .05$. ** $p < .01$.

General Discussion

The main purpose of this study was to establish whether a large discrepancy between the quality of video and photographic images affects face identification performance. According to theory, recognition of unfamiliar faces is very much dependent on image similarity. Therefore, better identification performance would be predicted when the image quality of matching stimuli are equalized. However, none of the three experiments produced results that

Table 7
Post Hoc ANOVAs for the Criterion Results in Experiments 1 and 3

Source	*df*	*MS*	*F*	Cohen's *f*
Recognition task				
Stimulus type (ST)	1	0.20	1.27	0.10
Congruity (C)	1	0.37	2.29	0.13
ST × C	1	0.00	0.00	0.00
Error	128	(0.16)		
Matching task				
ST	1	0.01	0.09	0.03
C	1	0.59	3.54	0.17
ST × C	1	0.86	5.15*	0.20
Error	128	(0.17)		

Note. Values enclosed in parentheses represent mean square errors. ANOVAs = analyses of variance.
* $p < .05$.

were consistent with this prediction. The data from the congruent and incongruent conditions did not differ significantly for the recognition task. The overall poor-recognition performance could have made the comparison of the two conditions difficult. This was particularly so for Experiment 3 in which the external features of the face were not shown. However, the results of the matching task did not have this problem. Although both conditions generated comparable sensitivity results when still images were used (Experiment 1), the advantage of the incongruent condition was present when dynamic video images were used (Experiment 2), or when external features of faces were removed (Experiment 3). Furthermore, these congruity advantages produced comparatively large effect sizes ($d > 0.7$ in both Experiments 2 and 3), whereas the nonsignificant effect in Experiment 1 showed no suggestion of an underlying large effect ($d < 0.2$).

Post hoc comparisons of the results from Experiments 1 and 2 show that inclusion of dynamic information did not produce an advantage over static stimuli. However, recognition performance in congruent and incongruent conditions was differentially affected by this information. Poorer performance was found in the congruent condition when dynamic rather than still images were used, whereas performance in the incongruent condition was unaffected by this change of stimulus type.

Not surprisingly, faces were more poorly identified when external features were removed (Experiment 3). This result is consistent with Bruce et al. (2001, Experiment 4) where matching performance was shown to be severely impaired when only internal features of faces were shown. The post hoc tests showed that d' measures were affected equally by the removal of external features.

Apart from sensitivity, there is also the issue of how congruent- and incongruent-image conditions affect a participant's criterion. It is quite conceivable that less similar images may prompt a more conservative criterion. However, the criterion data from the three experiments showed no evidence of this. The results differed only for the matching task in Experiment 1 in which the mean criterion of the congruent condition was unexpectedly more conservative than that of the incongruent condition. Overall, participants in both conditions shared a conservative bias in performing recognition and matching tasks. This was reflected by the consistently higher correct rejection rates relative to the hit rates. Perhaps because of the image difference between video and photographic stimuli, participants tended to perceive the target in one stimulus type as a different person from the one in the alternate stimulus type. However, criterion seems to be only modulated by some image differences because of lighting, pose, and so forth but not affected by other differences, such as resolution, size, and the like.

Overall, our results suggest that images differing in terms of resolution, size, color, dynamic information, and object spatial frequency create no greater deficits in identification performance than images that are made similar by equating image quality. It seems that if one of the matching images has poor resolution, identification is likely more reliable when the other image has high rather than equivalent resolution. The results thus demonstrate that face recognition is not only tolerant to a large discrepancy between image qualities of matching stimuli, it may actually benefit from rather than be impaired by a discrepancy in image quality. This is quite surprising because of prior evidence that recognition of unfamiliar faces is strongly influenced by image similarity. Our finding is not necessarily contradictory to the idea that image similarity affects unfamiliar face recognition. However, the generality of this theory needs to be constrained. Clearly, not all types of image similarity affect face recognition in the same manner. The present study shows the importance of specifying the kind of similarity that the visual system uses for unfamiliar face recognition.

The evidence from this study further suggests the necessity of investigating combined effects of image similarities because these effects may not be predicted from their individual effects. A combination of effects created by different types of similarity may not function in an additive manner. For example, although even a slight difference in spatial frequency overlap can affect face recognition (Liu et al., 2000), the effect of this incongruent spatial information vanishes when it is combined with other kinds of image difference such as lighting and pose change, as has been shown in this study.

The reason behind the advantage of incongruent over congruent, but poor-quality, images is not entirely clear. Because high-resolution photographs contain facial details that are not present in the poor-quality video, having these details or unmatched information should not facilitate matching performance. One interesting possibility, however, is that a high-resolution photograph provides the option for interpolating some of the missing information in the low-resolution video image, a function that the low-resolution photograph is lacking. Through this interpolation, the observer obtains or supplements more facial information related to the key internal features than the poor-quality video itself could offer.

A straightforward practical application of this study is that for criminal investigation, the best available quality of mug shots would be preferred when a criminal offender captured on low-quality video images is matched against an array of suspects in existing files, even though this means that identification has to be carried out with the least similar image-matching conditions. Our finding validates the methods used in prior applied studies (e.g., Bruce et al., 2001; Burton et al., 1999) in which data were collected under incongruent-image conditions. Also, it helps to clarify that the disadvantage of the low-quality video found in Henderson, Bruce, and Burton (2001) cannot be attributed to the incongruity between their low-quality video and the high-quality photograph because our data show that face recognition is robust against images with incongruent quality. Therefore, their results were more likely due to the quality of the image per se.

The results from the present study further suggest that dynamic video may be a worse choice than stills when the matching materials are both of similar poor quality. Finally, our results suggest that identification based on CCTV and photographic images may lead to relatively fewer false accusations, but at the expense of missing the targets. An awareness of this problem is therefore necessary to maximize the chance of determining the identity of the offender.

References

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society: Biological Sciences, 352,* 1203–1219.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5,* 339–360.

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied, 7,* 207–218.

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology, 77,* 305–327.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor quality video: Evidence from security surveillance. *Psychological Science, 10,* 243–248.

Campbell, R., Walker, J., & Baron-Cohen, S. (1995). The development of differential use of inner and outer face features in familiar face identification. *Journal of Experimental Child Psychology, 59,* 196–210.

Christie, F., & Bruce, V. (1998). The role of movement in the recognition of unfamiliar faces. *Memory & Cognition, 26,* 780–790.

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics, 58,* 602–612.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8,* 431–439.

Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception, 12,* 195–201.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4,* 330–337.

Harmon, L. D. (1971). The recognition of faces. *Scientific American, 229,* 70–82.

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology, 15,* 445–464.

Liu, C. H., & Chaudhuri, A. (1998). Are there qualitative differences between face processing in photographic positive and negative? *Perception, 27,* 1107–1122.

Liu, C. H., & Chaudhuri, A. (2000). Recognition of unfamiliar faces: Three kinds of effects. *Trends in Cognitive Sciences, 4,* 445–446.

Liu, C. H., & Chaudhuri, A. (2002). Reassessing the 3/4 view effect in face recognition. *Cognition, 83,* 31–48.

Liu, C. H., Collin, C. A., Rainville, S. J. M., & Chaudhuri, A. (2000). The effects of spatial frequency overlap on face recognition. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 956–979.

Nachson, I., Moscovitch, M., & Umilta, C. (1995). The contribution of external and internal features to the matching of unfamiliar faces. *Psychological Research, 58,* 31–37.

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences, 6,* 261–266.

Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Pike, G. E., Kemp, R. I., Towell, N. A., & Phillips, K. C. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition, 4,* 409–437.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception, 14,* 737–746.

Received September 17, 2002
Revision received December 9, 2002
Accepted December 13, 2002 ■